Business Research Methods:

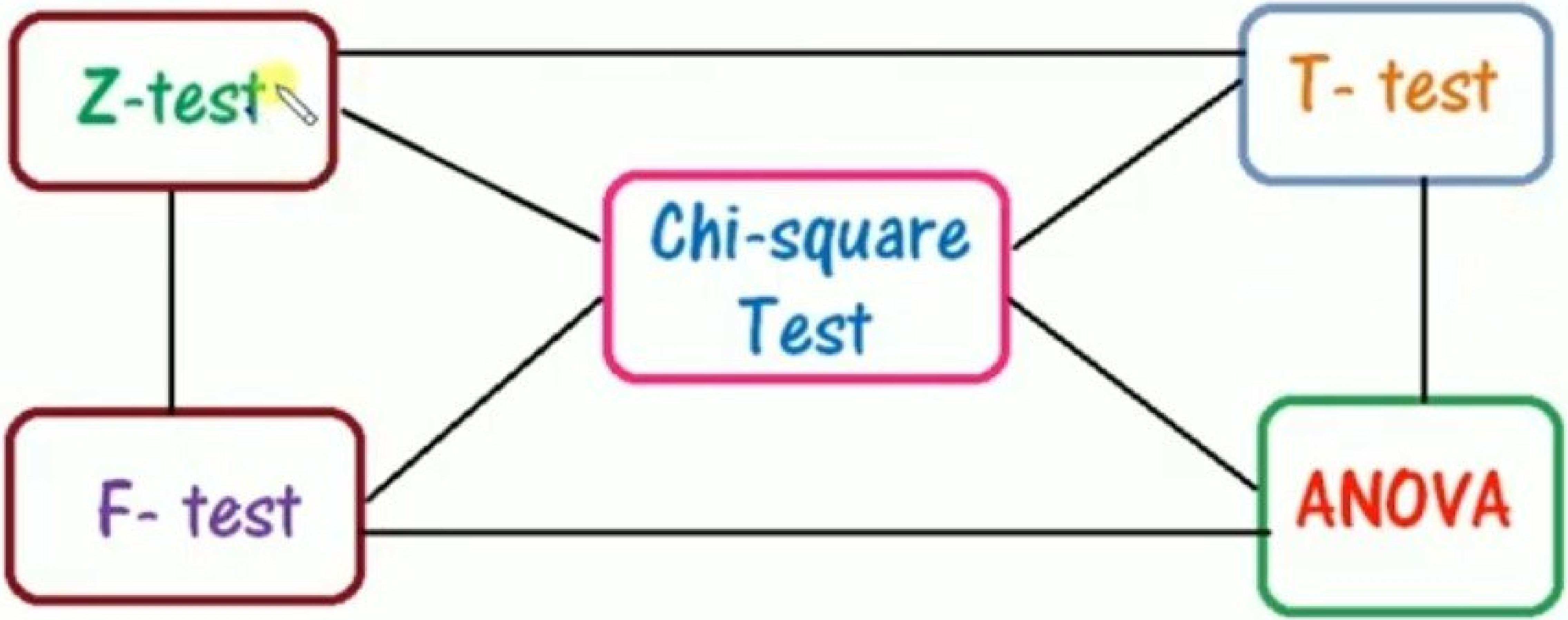
Non Parametric Analysis - Sign Test

By Dr. Satyabrata Dash

Professor- MBA Marketing

SMIT- PGCMS, Brahmapur

Testing of Hypothesis



Non-parametric Test

Many of the hypothesis tests require normal distributed populations or some tests require that population variances be equal.

What if, for a given test, such requirements cannot be met?

For these cases, statisticians have developed hypothesis tests that are "distribution free." Such tests are called nonparametric tests.

The several Non-Parametric tests are

- Sign Test
- Wilcoxon-Signed Rank test
- Mann-Whitney Test
- Kruskal-Wallis test
- Spearman's rank correlation coefficient

etc

Sign Test

It is the simplest of the entire non-parametric test.

As the name suggests, it is based on the signs (plus or minus) of the deviations rather than the exact magnitude of the variable values.

34 39 . . .

Single Sample Sign Test

It is used to test the hypothesis concerning the median for one population.

Suppose we want to test the hypothesis that median (η) of a population has a specified value, say η_0, i.e.,

$$H_0: \eta = \eta_0$$

Vs $\quad H_1: \eta \neq \eta_0$ (Two-tailed) $\qquad H_1: \eta > \eta_0$ (Right-tailed)

$$H_1: \eta < \eta_0 \text{ (Left-tailed)}$$

Procedure:

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the given population with median $\eta = \eta_0$ (under H_0).

Subtract, η_0 from each X_i's and write

1) Plus sign (+) if the deviation is positive.
2) Negative sign (-) if the deviation is Negative,
3) Zero (0) if the deviation is zero.

By the definition of Median, we have

$$P(X > Median) = P(X < Median) = \frac{1}{2}$$

Thus, under H_0 $(\eta = \eta_0)$:, we have

$$P(X > \eta_0) = P(X < \eta_0) = \frac{1}{2}$$

Hence, if H_0 is true, then the number of + signs should be approximately equal to the - signs.

If the difference in the number of plus (+) and minus (-) signs is due to chance variations (or fluctuations of sampling), then we fails to reject the H_0.

Notation:

After Discarding Zeros

T^{+} = Number of Positive Sign

T^{-} = Number of Negative Sign

$$T = \min(T^{+}, T^{-})$$

Example: For the null hypothesis, Median $(\eta) = 5$, compute the values of T^+, T^-, T for the following observations:

8 9 3 5 4 11

Solution: Null hypothesis: Median $(\eta) = 5$

Subtract 5 from each observations and writing the signs as

$+ \quad + \quad - \quad 0 \quad - \quad +$

Discard Zero and get

$T^+ =$ Number of positive signs

$= 3$

$T^- =$ Number of negative signs

$= 2$

Thus,

$T = \min(T^+, T^-)$

$= \min(3,2)$

$= 2$

Single Sample Sign test

(Small Samples, $n \leq 25$)

Procedure:

The following steps are summarized:

Step 1:
Set the Hypothesis:
H_0, H_1

Step 2:
compute T^+, T^-

Step 3:
Test statistics T

Step 4:
Critical region:

Procedure:

The following steps are summarized:

Step 1: Set the Hypothesis:

Null Hypothesis

$$H_0 : \eta = \eta_0$$

Alternative hypothesis

$H_1 : \eta < \eta_0$ (Left-tailed)

$H_1 : \eta > \eta_0$ (Right-tailed)

$H_1 : \eta \neq \eta_0$ (Two-tailed)

Step 2: compute T^+, T^-

Subtract η_0 from each observations

Discard zeros and hence compute T^+, T^- values.

Step 3: Test statistics:

$$T = \min\,(T^+, T^-)$$

Step 4: Critical region:

Define the critical region as $T \leq T_c$,

Where T_c is the critical value of T at given level of significance for one-tailed or two-tailed.

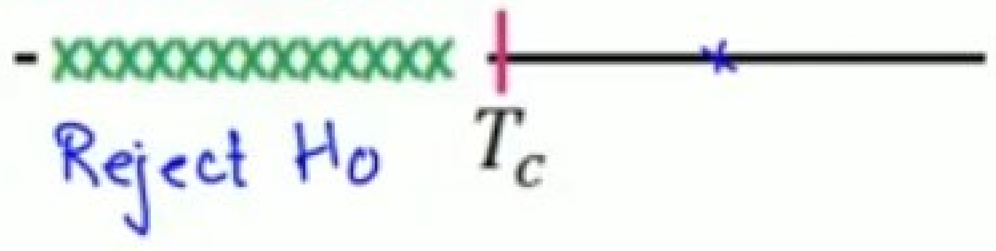
If calculated value (obtained from Step 3)

$$T \leq T_c$$

Then REJECT H_0, otherwise,

H_0 may be regarded as TRUE.

Critical Region



Example: Following are the responses to the question "How many hours do you study before a major Statistics test?"

6 5 1 2 2 5 7 5 3 7 4 7

Use the sign test to test the hypothesis at the 5% level of significance that the median number of hours a student studies before a test is 3. Given that the critical value of sign test for $n = 11$ at 5% level of significance for two-tailed test is 1.

Solution: Since the sample size is small $(n \leq 25)$.

Step 1: Null Hypothesis:

$$H_0 : \eta = 3$$

Alternative Hypothesis:

$$H_1 : \eta \neq 3 \text{ (Two-tailed)}$$

Step 2: Subtract 3 from each observation and writing the signs as

$$+ + - - - + + + 0 + + +$$

Discard Zero and get

$T^+ =$ Number of positive signs

$= 8$

$T^- =$ Number of negative signs

$= 3$

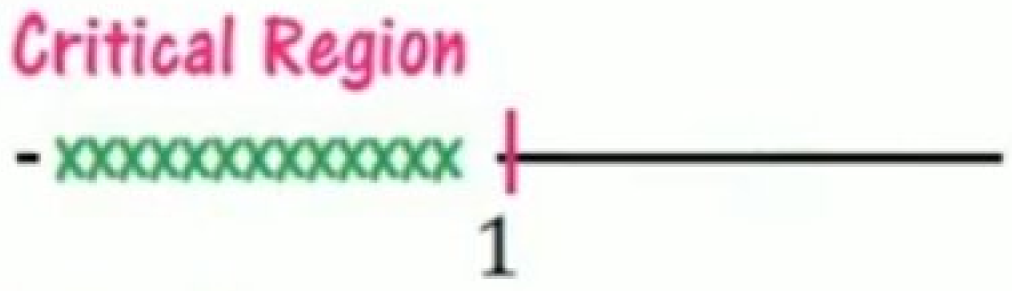
Step 3: Test statistics

$$T = \min(T^+, T^-)$$

$$= 3$$

Step 4: Critical Region

the critical value of sign test for $n = 11$ at 5% level of significance for two-tailed test is 1.



Thus, the critical region is $T \leq 1$

Since calculated value of $T = 3 > 1$,

so $\boxed{\text{we fail to REJECT } H_0.}$

Therefore, we conclude that the median number of an hour of study before a test is 3 hours.

Example: A teacher claims that the median time to do a particular type of Statistics problems is at most 3 minutes, but her students believe that the median time is more than 3 minutes. A random sample of 10 students completed the problem in the following times (in minutes)

2.5	2	4	4.5	4	2.5	4.5	3	3.5	5

Use the sign test with 5% level of significance to test the teacher's claim. Given that the critical value of sign test for $n = 9$ at 5% level of significance for one-tailed test is 1.

Solution: The sample size is small $(n \leq 25)$.

Step 1: Null Hypothesis:

$$H_0 : \eta \leq 3$$

Alternative Hypothesis:

$$H_1 : \eta > 3 \text{ (Right-tailed)}$$

Step 2: Subtract 3 from each observation and writing the signs as

$$- - + + + - + 0 + +$$

Discard Zero and get

$T^+ =$ Number of positive signs

$= 6$

$T^- =$ Number of negative signs

$= 3$

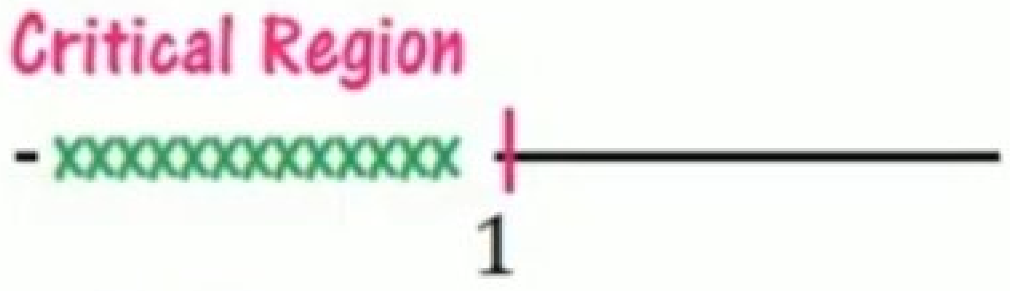
Step 3: Test statistics

$$T = \min(T^+, T^-)$$

$$= 3$$

Step 4: Critical Region

the critical value of sign test for $n = 9$ at 5% level of significance for two-tailed test is 1.



Thus, the critical region is $T \leq 1$

Since calculated value of $T = 3 > 1$,

so $\boxed{\text{we fail to REJECT } H_0.}$

Therefore, we conclude that the teacher claims that the Median time is at most 3 minutes MAY be regarded as TRUE.

Example: A travel agency, which promotes a particular holiday resort, advertises that the median cost per day of a motel in the city is \$50. A cautions traveller selects a random sample of 8 motels in that city and records the cost (in dollar) per day as follows:

52	51	49	50	53	52	48	47

Use the sign test, at 5% level of significance to test the travel agency's claim. Given that the critical value for $n = 7$ at given level for two-tailed test is 0.

Solution: We use Small Sample Sign test as the sample size is small $(n \leq 25)$.

Step 1: Null Hypothesis:

$$H_0 : \eta = 50$$

Alternative Hypothesis:

$$H_1 : \eta \neq 50 \text{ (Two-tailed)}$$

Step 2: Subtract 50 **from each observation and writing the signs as**

$$+ + - 0 + + - -$$

$T^+ = 4$; $T^- = 3$

Discard Zero and get

$T^+ =$ Number of positive signs

$= 4$

$T^- =$ Number of negative signs

$= 3$

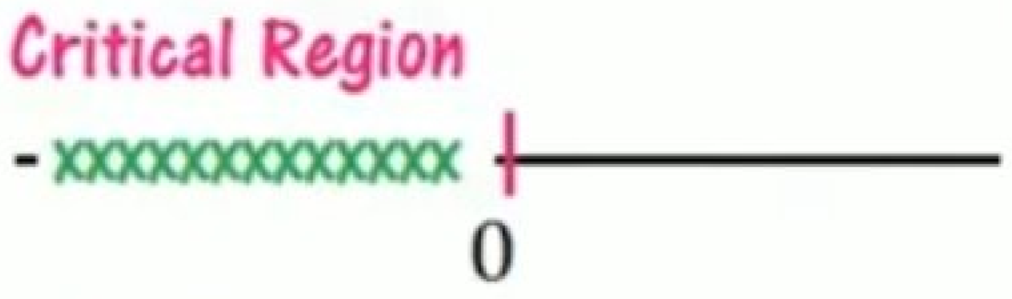
Step 3: Test statistics

$$T = \min(T^+, T^-)$$

$$= 3$$

Step 4: Critical Region

the critical value of sign test for $n = 7$ at 5% level of significance for two-tailed test is 0.



Thus, the critical region is $T \leq 0$

Since calculated value of $T = 3 > 0$,

so $\boxed{\text{we fail to REJECT } H_0.}$

Therefore, we conclude that the median cost per day of a city motel MAY not be different from \$50.

or

Single Sample Sign test

(Large Samples, $n > 25$)

Under $H_0 : P(T > \eta_0) = P(T < \eta_0) = 0.5$

$\Rightarrow$ $p =$ Probability of $+$ signs = 0.5,

which is constant for each trial.

Hence, under H_0, the variable T has Binomial distribution with parameter n and $p = 0.5$, i.e., $T \sim Bino(n, 0.5)$

Therefore, Under H_0:

$p = \frac{1}{2}$ $q = \frac{1}{2}$

$$E(T) = np$$
$$= \frac{n}{2}$$

$$\sigma = \sqrt{Var(T)}$$
$$= \sqrt{npq}$$
$$= \sqrt{\frac{n}{4}}$$
$$= \frac{1}{2}\sqrt{n}$$

If n is large, we use NORMAL distribution approximation to the binomial distribution, after applying the continuity correction.

Hence, for large samples, the test-statistics is

If $T^+ < \frac{n}{2}$

$$Z = \frac{(T^+ + 0.5) - \frac{n}{2}}{\frac{1}{2}\sqrt{n}} \sim N(0,1)$$

If $T^+ > \frac{n}{2}$

$$Z = \frac{(T^+ - 0.5) - \frac{n}{2}}{\frac{1}{2}\sqrt{n}} \sim N(0,1)$$

Procedure:

Step 1: Set the hypothesis H_0, H_1

Step 2: Compute T^+, T^-

Step 3: Compute the test statistics:

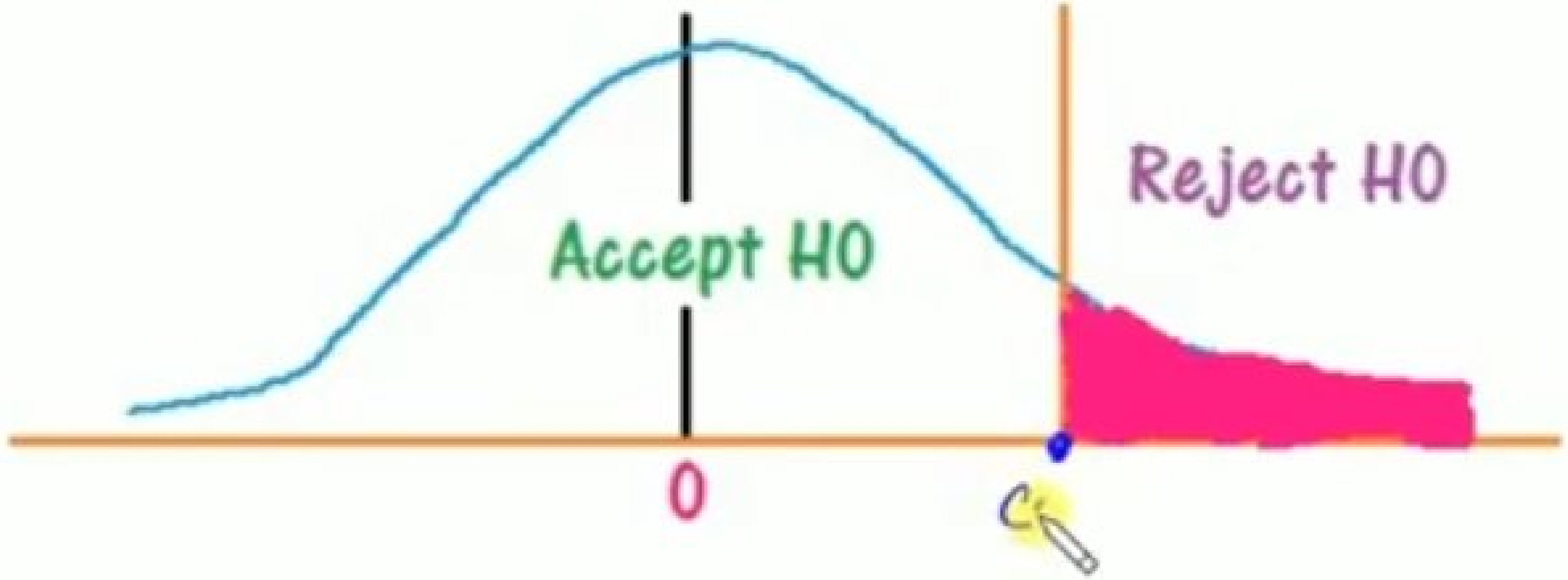
If $T^+ < n/2$

$$Z = \frac{(T^+ + 0.5) - \frac{n}{2}}{\frac{1}{2}\sqrt{n}}$$

If $T^+ > n/2$

$$Z = \frac{(T^+ - 0.5) - \frac{n}{2}}{\frac{1}{2}\sqrt{n}}$$

Step 4: Conclusion:



If computed $Z >$ critical value of Z at given level of significance, we REJECT H_0, otherwise we fail to reject H_0.

Example: To test the claim that the median age of mathematics faculty in the State community college at least 42 years, the results from a random sample of 32 mathematics faculty gave the following ages (in years)

56	62	61	54	52	32	24	35	50	42	52	49	26	31	31	54
38	36	45	53	37	40	38	31	29	25	45	32	49	39	36	38

Use the Sign test at the 5% level of significance to test the claim by (i) Critical value method (ii) p-value method.

Solution: Step 1:

Null Hypothesis:

$$H_0: \eta \geq 42$$

Alternative Hypothesis:

$$H_1: \eta < 42 \text{ (Left-tailed)}$$

Step 2: Subtract 42 from each observation and writing the signs as

$$+ \; + \; + \; + \; + \; - \; - \; - \; + \; 0 \; + \; + \; - \; - \; - \; +$$

$$- \; - \; + \; + \; - \; - \; - \; - \; - \; - \; + \; - \; + \; - \; - \; -$$

$T^+ =$ T^-

Discard Zero and get

$T^+ =$ Number of positive signs

$= 13$

$T^- =$ Number of negative signs

$= 18$

$$\boxed{\begin{aligned} n &= T^+ + T^- \\ &= 31 \end{aligned}}$$

Step 3: Test-Statistics:

Since $n > 25$, we use the Normal test as

As $T^+ < \frac{n}{2}$ so, we have

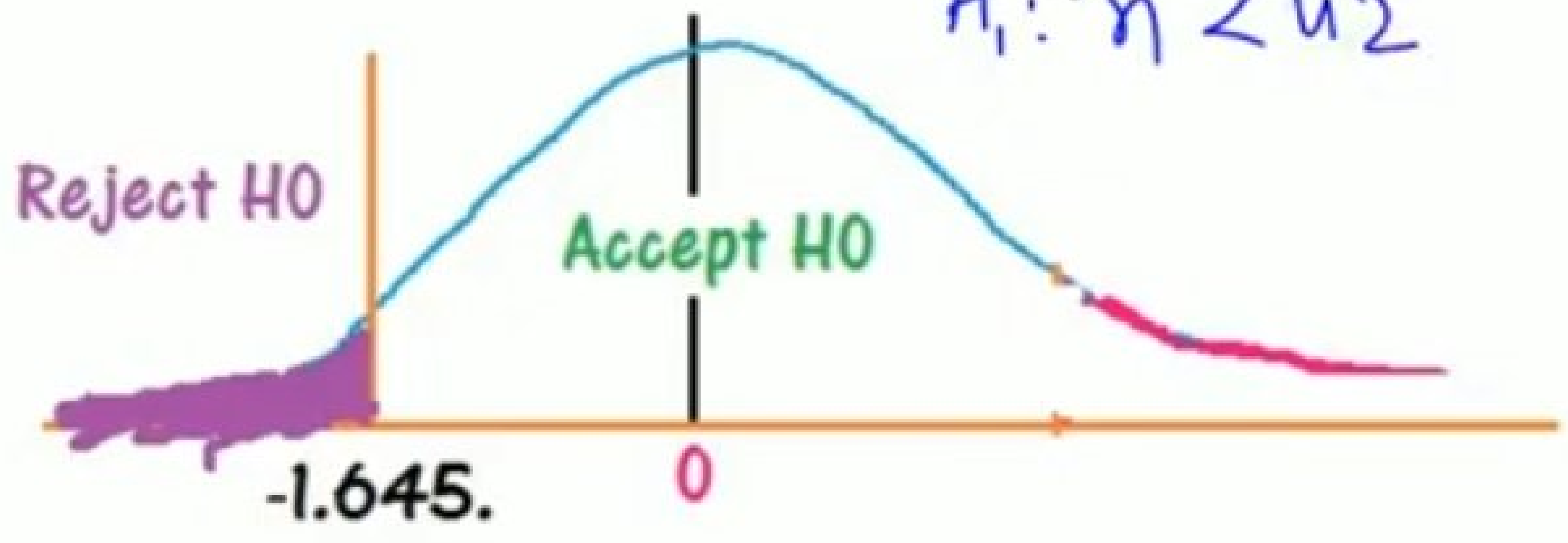
$$\boxed{\begin{aligned} Z &= \frac{(T^+ + 0.5) - \frac{n}{2}}{\frac{1}{2}\sqrt{n}} \\ &= \frac{(13+0.5)-\frac{31}{2}}{\frac{1}{2}\sqrt{31}} \\ &= -0.71842 \end{aligned}}$$

Step 4: The critical value of Z at 5% level for left-tailed is -1.645.

Since $-1.645 < -0.718$, so we fail to reject H_0.

$H_0: \eta \geq 42$

$H_1: \eta < 42$



So it is not sufficient evidence against $H_0: \eta \geq 42$. In other words, the claim that the median age of Mathematics faculty is at least 42 years is Valid.

Step 4: **By p-value method**:

$$p - \text{value} = P(Z < computed\ value\ of\ Z)$$

$$= P(Z < -0.71842)$$

$$= 0.23576$$

Since $\boxed{p - value > 0.05,}$

so we FAIL To reject H_0 at 5% level of significance.

So it is not sufficient evidence against $H_0: \eta \geq 42$. In other words, the claim that the median age of Mathematics faculty is at least 42 years is Valid.

Example: The results of a random sample of 35 long distance calls give the following durations (in minutes).

12	18	23	28	15	10	13	3	6	16	25	35	23	25	28	23	27	28
5	9	37	27	32	19	13	19	18	23	35	4	18	25	31	24	21	

Use the sign test at 1% level of significance to test the median length of long distance telephone calls is at most 15 minutes.

Solution: Step 1:

Null Hypothesis:

$$H_0: \eta \leq 15$$

Alternative Hypothesis:

$$H_1: \eta > 15 \text{ (Right-tailed)}$$

Step 2: Subtract 15 from each observation and writing the signs as

$$- + + + 0 - - - - + + + + + + + + +$$

$$- - + + + + - + + + + - + + + + +$$

Discard Zero and get

$T^+ =$ Number of positive signs

$= 25$

$T^- =$ Number of negative signs

$= 9$

$$\boxed{\begin{aligned} n &= T^+ + T^- \\ &= 34 \end{aligned}}$$

Step 3: Test-Statistics:

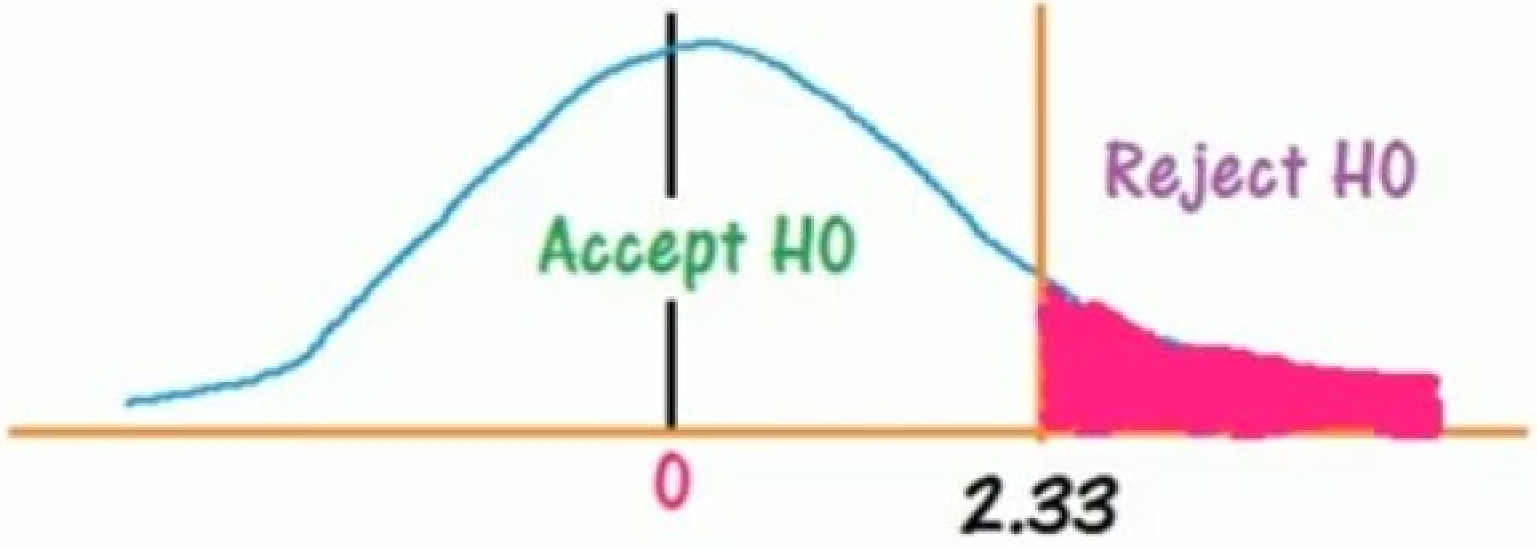
Since $n > 25$, we use the Normal test

As $T^+ > \frac{n}{2}$ so, we have

$$\boxed{\begin{aligned} Z &= \frac{(T^+ - 0.5) - \frac{n}{2}}{\frac{1}{2}\sqrt{n}} \\ &= \frac{(25-0.5)-\frac{34}{2}}{\frac{1}{2}\sqrt{34}} \\ &= 2.57 \end{aligned}}$$

Step 4: The critical value of Z at 1% level

for one-tailed is 2.33



As $2.57 > 2.33$, so we Reject H_0 .

Hence, we conclude that the median length of long distance telephone call is greater than 15 minutes.

$H_0 : \eta \leq 15$

$H_1 : \eta > 15$

(Right-tailed)

By p – value approach:

$H_0 : \eta \leq 15$

$H_1 : \eta > 15$

(Right-tailed)

$$p - value = P(Z > calculate\ value\ of\ Z)$$

$$= P(Z > 2.57)$$

$$= 0.0051$$

Since $p - value < 0.01$,

so reject H_0 at 1% level of significance.

Example: A random sample of 32 checking accounts at First state Bank gives the following monthly balances (in \$).

185	210	324	150	165	134	165	195	245	164
155	320	175	146	148	152	126	159	164	158
211	215	249	168	146	164	157	231	194	182
168	154								

Use the Sign test at 5% level of significance and test the hypothesis that the median monthly balance is at least \$200.

Solution: Step 1:

$H_0: \eta \geq 200$

$H_1: \eta < 200$ (left-tailed)

Step 2: Subtract 200 from each observation

−	+	+	−	−	−	−	−	+	−
−	+	−	−	−	−	−	−	−	−
+	+	+	−	−	−	−	+	−	−
−	−								

Discard Zero and get

$T^+ =$ Number of positive signs

$= 7$

$T^- =$ Number of negative signs

$= 25$

$$\boxed{\begin{aligned} n &= T^+ + T^- \\ &= 32 \end{aligned}}$$

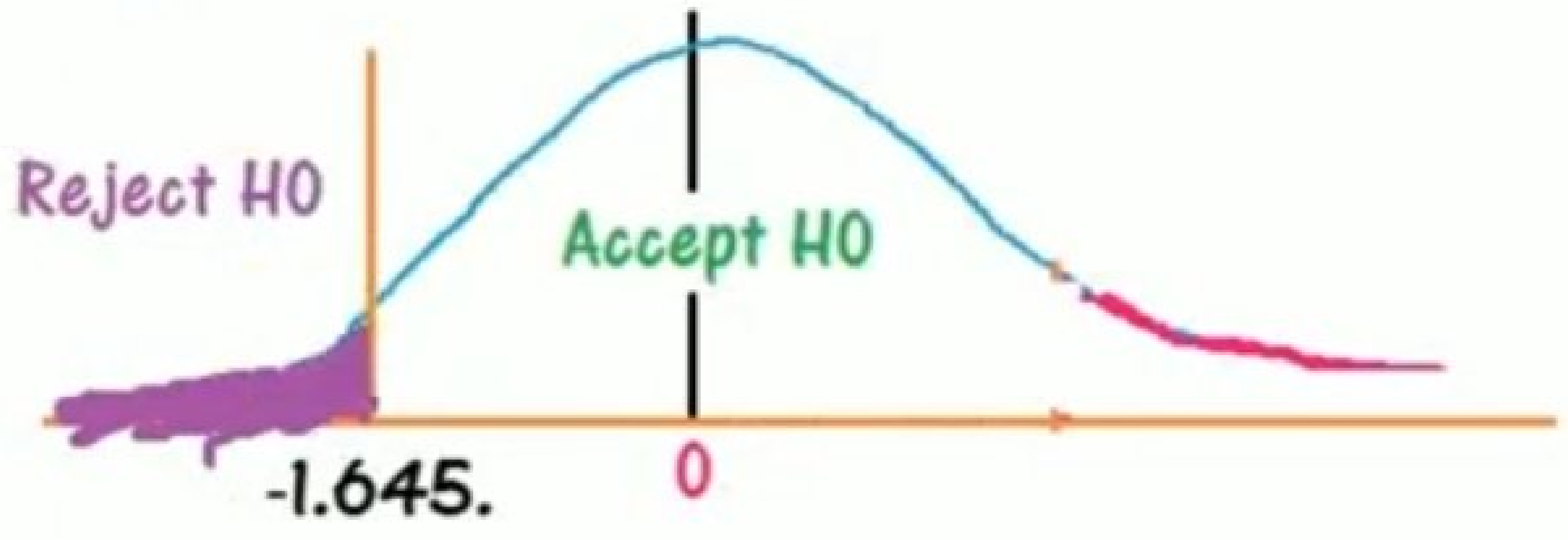
Step 3: Test-Statistics:

Since $n > 25$, we use the Normal test as

As $T^+ < \frac{n}{2}$ so, we have

$$\boxed{\begin{aligned} Z &= \frac{(T^+ + 0.5) - \frac{n}{2}}{\frac{1}{2}\sqrt{n}} \\ &= \frac{(7+0.5) - \frac{32}{2}}{\frac{1}{2}\sqrt{32}} \\ &= -3.0052 \end{aligned}}$$

Step 4: The critical value of Z at 5% level for left-tailed is -1.645.



As $-3.0052 < -1.645$,

so we Reject H_0 .

Hence, we conclude that the median monthly balance is less than $200.

$H_0: \eta \geq 200$

$H_1: \eta < 200$ (left-tailed)

$H_0: \eta \geq 200$

$H_1: \eta < 200$ (left-tailed)

By p - value approach:

$$p - value = P(Z < calculate\ value\ of\ Z)$$

$$= P(Z < -3.0052)$$

$$= 0.00001$$

Since $p - value < 0.05$,

so reject H_0 at 5% level of significance.

Two-Sample Sign Test

Informal spoken	Formal written
5	5
4	2
5	3
4	4
3	1
2	3
4	3
5	1
4	2
2	3
4	2
4	3
5	3
3	5
3	0

X	Y	X-Y	(+/-)
5	5	0	0
4	2	2	(+)
5	3	2	(+)
4	4	0	0
3	1	2	(+)
2	3	-1	(-)
4	3	1	(+)
5	1	4	(+)
4	2	2	(+)
2	3	-1	(-)
4	2	2	(+)
4	3	1	(+)
5	3	2	(+)
3	5	-2	(-)
3	0	3	(+)

Success	Failure
10	3

Z= (x-npo)/Root of ((npo*(1-po))

x- npo = 3.5

npo*(1-po)= 3.25

Root of ((npo*(1-po))= 1.80

Ho: p = 1/2

Ha: p> 1/2

X= 10

n= 13

P= 1/2

Z= 1.94

Absolute Z = !Z!= 1.94

Tabulated value of Z at ά = 0.05= 1.645

Calculated value > tabulated value

H0 is rejected differ significantly

43